Harnessing the power of individual participant data in a meta-analysis of the benefits and harms of the Incredible Years parenting program

Leijten, Patty; Gardner, Frances; Landau, Sabine; Harris, Victoria; Mann, Joanna; Hutchings, Judith; Beecham, Jennifer; Bonin, Eva-Maria

Journal of Child Psychology and Psychiatry

DOI:
10.1111/jcpp.12781

Published: 01/02/2018

Peer reviewed version

MANUSCRIPT ACCEPTED FOR PUBLICATION IN *JOURNAL OF CHILD PSYCHOLOGY AND PSYCHIATRY (JCPP)*, 30 MAY 2017

Harnessing the Power of Individual Participant Data in a Meta-Analysis of the Benefits and Harms of the Incredible Years Parenting Program

Patty Leijten (University of Oxford, University of Amsterdam)

Frances Gardner (University of Oxford)

Sabine Landau & Victoria Harris (King's College London)

Joanna Mann (University of Oxford)

Judy Hutchings (Bangor University)

Jennifer Beecham & Eva-Maria Bonin (Kent University, London School of Economics)

Stephen Scott (King's College London)

Conflicts of Interest

PL, FG, SL, JH, & SS were involved in evaluations of the Incredible Years parenting program included in this study. JH reports personal fees for the delivery of leader training for Incredible Years. None of the other authors have conflicts of interests to declare.

Abstract

Background: Parenting programs aim to reduce children's conduct problems through improvement of family dynamics. To date, research on the precise benefits and possible harms of parenting programs on family well-being has been unsystematic and likely to be subject to selective outcome reporting and publication bias. Better understanding of program benefits and harms requires full disclosure by researchers of all included measures, and large enough numbers of participants to be able to detect small effects and estimate them precisely. **Methods:** We obtained individual participant data for 14 out of 15 randomized controlled trials on the Incredible Years parenting program in Europe (total $N = 1799$). We used multilevel modeling to estimate program effects on thirteen parent-reported outcomes, including parenting practices, children's mental health, and parental mental health. **Results:** Parental use of praise, corporal punishment, threats and shouting improved, whilst parental use of tangible rewards, monitoring, or laxness did not. Children's conduct problems and ADHD symptoms improved, whilst emotional problems did not. Parental mental health (depressive symptoms, self-efficacy, stress) did not improve. There was no evidence of harmful effects. **Conclusions:** The Incredible Years parenting program improves the aspects of family well-being that it is primarily designed to improve: parenting and children's conduct problems. It also improves parent-reported ADHD symptoms in children. Wider benefits are limited: the program does not improve children's emotional problems or parental mental health. There are no signs of harm on any of the target outcomes.

Keywords: parenting program; Incredible Years; conduct problems; individual participant data (IPD) meta-analysis.

Abbreviations: ADHD = Attention Deficit Hyperactivity Disorder.

Parenting programs are the key strategy for the prevention and early treatment of children's conduct problems (Weisz & Kazdin, 2010). Most established programs are based on social learning theory, and encourage parents to adopt behaviors to improve their relationship with their child, reinforcement techniques to increase positive child behavior, and non-violent discipline techniques to reduce children's conduct problems (Kaehler, Jacobs, & Jones, 2016). Although the effects of parenting programs on children's conduct problems have been widely studied, research on benefits and possible harms of parenting programs for family well-being more broadly has been unsystematic and likely to be subject to selective outcome reporting and publication bias (Dwan et al., 2008).

It is often unclear why some trials, and not others, include particular secondary outcome measures (e.g., parental depression), as only a minority of the trial protocols are registered (Cybulski, Mayo-Wilson, & Grant, 2016). Similarly, potential harmful effects are almost never studied in parenting program research, while it is well-established that social interventions can sometimes cause harm (e.g., Bonell, Jamal, Melendez-Torres, & Cummins, 2015). Better understanding of a program's benefits and harms requires full disclosure by researchers of all included measures, and large enough numbers of participants to be able to detect small effects and estimate them precisely. For these reasons we conducted an individual participant data (IPD) meta-analysis of the effects of one of the main parenting programs recommend by influential guidelines (e.g., Blueprints, NICE) for the prevention and treatment of children's conduct problems: The Incredible Years parenting program (Webster-Stratton & Reid, 2010).

Incredible Years is largely similar in content to many other established parenting programs (Kaehler, Jacobs, & Jones, 2016; Figure 1), although its strong emphasis on a collaborative group approach differs from most other parenting programs. The program has become part of usual care in several countries and shows robust effects in independent

evaluations across countries (Leijten, Melendez-Torres, Knerr, & Gardner, 2016). We considered the following outcomes as potentially beneficially or adversely affected by the program:

Parenting Practices

Improved parenting behavior is the most proximal target of Incredible Years, but the specific aspects of parenting that change, and those that do not change, are unclear. Harsh and inconsistent parenting and a lack of warmth and involvement affect child well-being beyond conduct problems. Even mildly harsh parenting is associated with harmful biological effects on children, such as dysfunctional cortisol secretion patterns (e.g., Scott, 2012).

Children's Mental Health

ADHD is the most prevalent comorbid problem to children's conduct problems. Despite their different correlates, with ADHD being more related to cognitive deficits and conduct problems to psychosocial factors, both have been associated with more negative parent-child interactions (Hinshaw, 1992). Some studies find that Incredible Years reduces children's ADHD symptoms (e.g., Jones, Daley, Hutchings, Bywater, & Eames, 2007). Others, however, were unable to replicate this (e.g., Leijten, Raaijmakers, Orobio de Castro, van den Ban, & Matthys, 2015).

Symptoms of anxiety and depression are common in children with conduct problems (Levy, Hawes, & Johns, 2015). The extent to which these symptoms can be altered by changing the same parenting behaviors that alter conduct problems is unclear. Some studies suggest that Incredible Years reduces emotional problems (e.g., Webster-Stratton & Herman, 2008). Others failed to replicate these findings (e.g., Axberg & Broberg, 2012) or did not report on emotional problems (e.g., Gross et al., 2009).

Parental Mental Health

Parental mental health problems are particularly prevalent in families with child conduct problems (McLaughlin et al., 2012). Traditional systematic reviews suggest that parenting programs reduce parental depressive symptoms, stress, and improve self-efficacy (Barlow, Smailagic, Huband, Roloff, & Bennett, 2014; Furlong et al., 2012). These reviews, however, cannot account for selective reporting and publication bias.

Possible Harmful Effects

Harmful effects are almost never studied in parenting program research. Parents rarely report potentially harmful outcomes in qualitative studies, but it does happen (e.g., Furlong & McGilloway, 2012). A systematic review (Furlong et al., 2012) planned to examine burden on families participating in a parenting program, but found no studies reporting these outcomes. Harmful effects of psychosocial interventions can be adverse effects on expected outcomes, or adverse effects on unexpected outcomes (Bonell et al., 2015).

Objectives

The objectives of this study were to test the benefits and possible harms of the Incredible Years parenting program on three domains of family wellbeing: parenting (i.e., positive and negative), children's mental health (i.e., conduct problems, ADHD symptoms, emotional problems), and parental mental health (i.e., depressive symptoms, stress, self-efficacy). We chose to focus on evaluations of Incredible Years that were (i) independent of the program developer, and (ii) in European countries, to ensure sufficient homogeneity in the usual services that children receive across trials, and because some family programs developed outside of Europe have not worked well in Europe (e.g., Robling et al., 2016).

Methods

Protocol, Registration, and Reporting

The study protocol is available on the National Institute of Health Research (NIHR) Public Health Research website *[blinded for review]* and was approved by the Departmental

Research Ethics Committee of *[blinded for review]*. We followed PRISMA-IPD guidelines for reporting meta-analyses of individual participant data (Stewart et al., 2015).

Eligibility Criteria

We sought to include all completed randomized trials of the Incredible Years parenting program for children aged 1-12 years conducted by independent researchers in Europe. We excluded trials, or conditions within trials, that (1) were not randomized; (2) included non-parenting programs, such as the Incredible Years child program; or (3) evaluated abbreviated, non-standard versions of the original 12–14 session Incredible Years program.

Identifying and Selecting Trials

Trials were identified through (1) systematic searches in five databases (CINAHL, Embase, Global Health, MEDLINE, and PsycINFO), (2) the Incredible Years website overview of trials, (3) the European Incredible Years mentors' network, and (4) asking experts. See Figure 2 for our search results. Embase, Global Health, MEDLINE, and PsycINFO were searched via OVID using the following search terms: 1. incredible years.mp; 2. webster-stratton.mp; 3. 1 or 2. CINAHL was searched via EBSCO using the phrase "incredible years". Searches in January 2015 revealed no further completed trials. Eligibility was assessed by the second author and double checked by four additional authors. There were no differences of opinion.

Data Collection Process

Anonymized data were requested for fifteen trials. Data for fourteen trials were supplied; data for one trial (Patterson et al., 2002) had no longer been retained by the investigators. Raw (i.e., unscored) individual item-level data were supplied in SPSS and checked for missing items, internal consistency, and consistency with trial protocols and reports. Most trials included data from only one parent, therefore data on the parent who was

the primary caregiver (98% mothers) were used. Copies of the original questionnaires were supplied to check for consistent use across trials. There were very limited data available from two measures (teacher-reported conduct problems and parent-reported marital quality; less than 400 of 1799 families, or less than four trials) and these were therefore excluded.

Most trials included a measure of observed parenting and child behavior, but the complex and diverse nature of these observational data made harmonization difficult and potentially hazardous. Specifically: (1) data were on different scales, such as frequency counts of discrete parenting behaviors versus global ratings of parental affect; (2) numbers of trials that used similar coding schemes were small; (3) key assumptions could not be tested (e.g., whether free play tasks in the lab elicit the same parenting behavior as tidy-up tasks in the home). Thus, it would not possible to establish the validity of any harmonization strategy, because correlations between observational measures are neither available from previous work, nor was there overlap in observational measures within trials to empirically assess such correlations.

None of the trials explicitly focused on harmful effects. The data therefore allow to study some aspects of harm (i.e., adverse effects on outcomes related to the program aims), but not other aspects (i.e., adverse effects on outcomes other than those related to the program aims) (Bonell et al., 2015).

Overview of Trials

Table 1 provides an overview of the fourteen included trials. Trials were conducted in England (k = 6), Wales (k = 2), the Netherlands (k = 2), Ireland (k = 1), Norway (k = 1), Portugal (k = 1), and Sweden (k = 1). Ten trials were indicated prevention or treatment trials that included children screened for conduct problems. Seventy per cent of the children in these trials scored above the clinical cut-off on the Eyberg Child Behavior Inventory (ECBI; Eyberg & Ross, 1978). Four trials were selective prevention trials that targeted high risk

families (e.g., socioeconomically disadvantaged families or mothers released from incarceration). Thirty per cent of the children in these trials scored above the clinical cut-off on the ECBI.

Outcome Measures

Outcomes were assessed immediately after the program ended, around four to six months after baseline assessment. To allow data synthesis across trials, data from different measures had to be harmonized. Data across measures on children's mental health and parental depression were harmonized using norm deviation scores (i.e., converting raw scale scores on one measure to norm deviation scores of this measure, and converting these norm deviation scores in turn to raw scale scores on another measure). Data across measures of parenting behavior were harmonized by selecting items that theoretically fitted the same constructs. Details of the data harmonization strategies, available data on each outcome contributed by each trial, and evidence for the reliability and validity of our harmonization approach are explained in the Online Supplemental Material.

Parenting practices

Negative parenting. Three established measures were used across trials: Parenting Practices Inventory (PPI, four trials; Conduct Problems Prevention Research Group, 1996); Alabama Parenting Questionnaire (APQ, two trials; Shelton, Frick, & Wootton, 1996); and Parenting Scale (PS; four trials; Arnold, O'Leary, Wolff, & Acker, 1993). Four trials used an interview format designed specifically for the study. Three trials included multiple parenting measures. For these trials, we chose (1) measures that contributed data to more outcomes (e.g., on both positive and negative parenting) over measures that provided data to fewer outcomes (e.g., PPI data were preferred over PS data), and (2) well-validated measures over measures developed by the study authors (e.g., APQ data were preferred over interview data). Across measures, items fitted four theoretical constructs of negative parenting: *Corporal*

punishment, defined as any physical punishment; *threatening*, defined as threatening the child to punish him/her (but not directly punishing him/her); *laxness*, defined as intending to discipline the child, but not actually following it through; and *shouting*, defined as raising of the voice, shouting, scolding, and use of mean or bad language. Between nine and ten trials contributed data to each of the constructs.

Positive parenting. The same measures were used as for self-reported negative parenting. Items fitted three theoretical constructs of positive parenting: *Praise*, defined as giving verbal compliments in response to the child's behavior; *tangible rewards*, defined as non-verbal rewards, such as privileges or stickers on a chart; and *monitoring*, defined as parental supervision and knowledge of the child's whereabouts when the child was out of the parents' sight, including knowing the child's friends. Between six and nine trials contributed data to each of the constructs.

Children's mental health

Children's conduct problems. The ECBI Intensity Scale (Eyberg & Ross, 1978) was used to assess parent-reported children's conduct problems. For two trials scores on the Parental Account of Children's Symptoms (PACS; Taylor, Schachar, Thorley, & Wieselberg, 1986) were converted into ECBI scores using norm deviation scores. PACS and ECBI scores correlated $r = .71$ in our sample, based on data from four trials that included both measures. Internal consistency of the ECBI and PACS ranged from $\alpha = .79$ to $\alpha = .95$. One trial did not contribute data because of the young age of the children.

ADHD symptoms. The Hyperactivity scale of the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) was used to assess parent-reported ADHD symptoms. For two trials scores from the Child Behavior Checklist (CBCL; Achenbach, 1991) were converted to SDQ scores using norm deviation scores; for one trial scores from the PACS were converted to SDQ scores. Thirteen trials contributed data.

Emotional Problems. The Emotional Problems scale of the SDQ was used to assess parental report of children's emotional problems. For two trials scores from the CBCL were converted to SDQ scores using norm deviation scores; for one trial scores from the PACS were converted to SDQ scores. Twelve trials contributed data.

Parental Mental Health

Parental depressive symptoms. The Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) was used to assess parental depressive symptoms. For one trial scores from the Brief Symptom Inventory—depression subscale (Derogatis & Spencer, 1982) were converted to BDI scores; for one trial scores from the General Health Questionnaire (Williams & Goldberg, 1988) were converted to BDI scores. Internal consistency ranged from $\alpha = .87$ to $\alpha = .93$. Twelve trials contributed data.

Parenting stress. The Parenting Stress Index—Short Form (Abidin, 1995) was used to assess parental stress. Internal consistency was $\alpha = .95$. Six trials contributed data.

Parental feelings of self-efficacy. The Parental Sense of Competence Scale (Johnston & Marsh, 1989) was used to assess feelings of parental self-efficacy. Internal consistency was $\alpha = .84$. Five trials contributed data.

Risk of Bias of Individual Trials

We assessed risk of bias in the included trials (as high, low or unclear) using the Cochrane tool (Figure 3).

Risk of Bias Across Trials

Risk of bias was low with regard to heterogeneity of the program, because Incredible Years is strictly protocolled, including training and supervision of the staff, and is therefore expected to be very similar across trials. Risk of bias may be higher for some data harmonization procedures. For some constructs (e.g., self-reported parenting practices) data came from up to four different measures. Although mostly well-validated measures were

used, we were not always able to test the extent to which measures indeed measured the same construct. Relatedly, several assumptions had to be made. For example, norm scores were often not available for all measures from all countries. We therefore used the same norm scores for all samples.

Statistical Methods

Analyses were conducted in the pooled dataset of harmonized data from all trials in Stata v.14. Three statistical issues needed addressing: (i) the pooled data have a hierarchical structure with families (Level 1) nested within Incredible Years therapy groups (Level 2) within the intervention condition, and therapy groups nested within trials (Level 3); (ii) various design features of the original trials, such as cluster randomization, stratified randomization, and changes in allocation ratios over the duration of the trial; and (iii) we needed to minimize missing data biases due to large amounts of missing data for some of the pooled outcome variables. We addressed these issues by using multilevel modeling (random effect modeling) to capture the hierarchical structure of each variable. We accommodated design features by trial-specific fixed effects, and fitted resulting models by maximum likelihood which is valid under a missing at random (MAR) assumption about the missing data.

We ran two sets of analyses, each of which made some further assumptions: (A) univariate analyses that fully captured the hierarchical structure but required more restrictive MAR assumption and (B) a simplified multivariate analysis that dealt with all the outcome variables in one model and estimates the effects on each outcome in the presence of missing data on other outcomes, but did not account for all the trial features. The univariate models included the following variables: (i) covariates for the baseline values of the outcome of interest; (ii) fixed effects for the intervention condition and trial design features; (iii) random effects for the trial level; (iv) random effects for therapy group; (v) dummy variables for each

outcome and interactions between these dummy variables and covariates; (vi) scores on the measure of children's conduct problems, the primary outcome of the trials, to allow for further relaxation of the missing data assumptions. The multivariate model included similar variables, but left out the random effects for therapy group. We estimated the model with an unstructured covariance structure, which allows correlations to vary between each pair of outcomes. We estimated the model with an unstructured covariance structure, which allows correlations to vary between each pair of outcomes. Heterogeneity estimates (I^2) between studies were derived by first estimating the effect size within each trial and the associated standard errors in a univariate model, and then analyzing these trial specific effect sizes in the same way as would be done in a traditional meta-analysis using the Stata command "metan."

Results

Pooled Sample

Table 2 shows the baseline scores of families in the intervention and control condition. Boys showed higher levels of conduct problems (M = 141.44) and ADHD symptoms (M = 6.07) than girls (M = 131.75 and M = 5.36, respectively). No such pattern existed for emotional problems (M = 3.34 for boys and M = 3.38 for girls). Children's mental health problems did not correlate with children's age (rs ranged -.05 to .03). Conduct problems correlated moderately with ADHD symptoms (r = .41). Both conduct problems and ADHD symptoms correlated weakly with emotional problems (both r = .26). There was substantial variation across trials in how many parents scored above the clinical cut-off for depression (2% to 48%). In the pooled sample 23% of the parents reported clinical levels of depressive symptoms.

IPD Meta-Analysis

We report the results of the univariate analyses. Univariate and multivariate analyses results did not differ substantially (see Table 2 and Figure 4). Where results did differ we

report both results. Standardized effect sizes β express intervention effects in units of baseline standard deviations. For negative variables (i.e., those for which a lower value is a desired outcome) a negative coefficient reflects benefit, while a positive coefficient reflects harm.

Parenting practices

Incredible Years reduced the majority of harsh and inconsistent parenting constructs: corporal punishment (β =-.22, 95% CI -.42 to -.01), threatening (β = -.21, 95% CI -.36 to -.06), and shouting (β = -.31, 95% CI -.61 to -.01). The reduction in parental laxness (β = -.15, 95% CI -.37 to .07) was not statistically significant in the univariate analysis, but did reach statistical significance in the multivariate approach.

The program improved one aspect of positive parenting: parents who had received the program reported praising their child more frequently (β =.26, 95% CI .01 to .51). It did not improve parental use of tangible rewards (β = .15, 95% CI -.16 to .45) or monitoring (β = .05, 95% CI -.08 to .18).

Children's conduct problems

Incredible Years reduced parent-reported conduct problems (β = -.35, 95% CI -.51 to -.19). In the intervention condition, the percentage of children who scored above the 90th percentile reduced from 44% to 18% – a difference of 26% of the total. In the control condition, this percentage reduced from 37% to 29% – a difference of 8% of the total.

The program also reduced children´s ADHD symptoms (β = -.30, 95% CI -.44 to -.17). In the intervention condition, the percentage of children who scored above the borderline threshold on ADHD symptoms reduced from 54% to 42% – a difference of 12% of the total. In the control condition, this percentage reduced from 52% to 50% – only a difference of 2% of the total.

The program did not reduce children's emotional problems (β = -.06, 95% CI -.18 to .06). In the intervention condition, the percentage of children who scored above the

borderline threshold on emotional problems reduced from 40% to 32% – a difference of 8% of the total. The same reduction was found in the control condition.

Parental mental health

Incredible Years did not improve parents' feelings of depression, stress, or self-efficacy. There was a trend towards a reduction in parental depressive symptoms (β = -.08, 95% CI -.17 to .01), but this did not yield significance. At baseline, 25% of the parents in the intervention condition and 20% of the parents in the control condition scored above the cut-off score for depression. At post test, this was 14% of the parents in both conditions. Similarly, the program did not improve parenting stress (β = -.18, 95% CI -.44 to .07) or feelings of parental self-efficacy (β = -.32, 95% CI -.77 to .13).

Discussion

We synthesized individual participant data from fourteen independent randomized trials to estimate the effects of the Incredible Years parenting program on family well-being. In addition to reducing children's conduct problems, the program reduced parent-reported ADHD symptoms, parental use of corporal punishment, threats and shouting, and it increased parents' use of praise. The program did not reduce children's emotional problems or parental use of tangible rewards or monitoring. There was also no evidence of it improving parental mental health (depressive symptoms, parenting stress, or self-efficacy). Estimation of the effects of the program on parental laxness was sensitive to the analytic method used. We therefore conclude that this effect is unclear at this point and needs to be studied further. There were no suggestions of harm.

Only a minority of the trial protocols are registered (Cybulski et al., 2016) and selective outcome reporting and publication bias in reporting on the effects of psychosocial interventions seem common (Dwan et al., 2008). Our study is the first to overcome selective outcome reporting and publication biases in parenting intervention research by including data

from all outcome measures of fourteen trials of the same program. Our use of individual participant data allowed for increased statistical power and precision to estimate the sizes of benefits and harms. Moreover, our use of data from multiple contexts and seven countries increased the generalizability of our findings to a broader population.

Many trials and meta-analyses rely on broad constructs of parenting behavior, such as "positive parenting" and "negative parenting" (e.g., Furlong et al., 2012). Our finding that Incredible Years improved several, but not all aspects of positive and negative parenting indicates the importance of disentangling broad constructs into more specific behaviors, to better understand the precise impact of interventions on parenting behavior.

Incredible Years reduced parent-reported ADHD symptoms in children. Several official guidelines recommend parenting programs as the first line intervention for children's ADHD symptoms (e.g., National Collaborating Centre for Mental Health, 2009). Our study was not designed to test the suitability of the Incredible Years program for children at risk for ADHD, but showed that Incredible Years has additional effects on ADHD symptoms in children with conduct problems.

We found that Incredible Years reduced children's externalizing behavior specifically (i.e., conduct problems and ADHD symptoms), and not children's internalizing behavior (i.e., emotional problems). Emotional problems in our pooled data sample were relatively prevalent and statistical power allowed the detection of very small effect sizes. Thus, we were well-equipped to detect any effects of Incredible Years on children's internalizing behavior. Our finding supports dominant theories about internalizing child behavior in which children's social withdrawal, cognitive distortions, and maladaptive coping strategies are more central than coercive parenting (e.g., Ginsburg & Schlossberg, 2002).

Our finding that Incredible Years did not improve parental mental health contradicts findings from recent systematic reviews and meta-analyses (Barlow et al., 2014; Furlong et

al., 2012). Traditional meta-analysis might be subject to selective outcome reporting and publication bias (Littell, Corcoran, & Pillai, 2008). Most trials that report effects of parenting programs on parental mental health report statistically significant effects. This might lead to an overestimation of the overall effects of parenting programs on parental mental health. The extent to which parenting programs affect parental mental health might depend on program or sample characteristics, such as baseline depression rates. These rates varied substantially between trials in our sample (2% to 48%) and moderator analyses would be needed to identify whether program effects vary between subgroups of families.

There were no signs of harmful effects. If there were any adverse main effects, these would most likely be picked up by our large sample size. The outcomes included in the primary trials were most likely included based on the hypothesis that they might reveal benefits, rather than harms. Revealing harmful effects sometimes requires different outcome measures (Bonell et al., 2015; Scott & Young, 2016). Thus, our study was equipped to pick up on some aspects of harm (i.e., adverse effects on outcomes related to the program aims), but not other aspects (i.e., adverse effects on outcomes other than those related to the program aims). To complicate matters even further, harmful effects may arise in particular subsets of families. For example, feelings of self-efficacy might decrease in parents whose children fail to benefit from the program. It was beyond the scope of our study to include these subgroup analyses, for which theory is scarce.

Strengths and Limitations

Our unique IPD meta-analysis yields more transparency and a much larger sample from multiple contexts and countries than any other study in this field, and findings that are potentially well-generalizable. Yet, several limitations merit attention. First, we relied on parent-reported outcome measures. Parent reports can be biased because they are not blind to intervention conditions (Sonuga-Barke et al., 2013). Our ability to include measures

depended upon inclusion and possibilities for harmonization of measures in the individual trials. Data on teacher-reported child behavior were available for only a few trials and data on observed child and parenting behavior were too diverse in terms of content (e.g., child compliance versus negative affect) and structure (e.g., Likert scales versus frequency counts) for harmonization within the scope of this study. A recent meta-analysis is reassuring in showing that the effects of Incredible Years on children's conduct problems were at least as strong when based on independent observations compared to parent-report (Menting, Orobio de Castro, & Matthys, 2013). Second, several assumptions had to be made in harmonizing data. Norm deviation scores and item-matching approaches rely on the assumption that instruments measure the same construct with the same measurement error. Third, because most trials used a wait list control condition, longer-term data were often not available since the wait list families went on to be treated. We therefore only examined immediate program effects.

Implications

The extent to which programs designed for one aspect of child mental health have a broader impact on family well-being has important clinical implications. Our findings do not suggest that the Incredible Years parenting program should be the target intervention for alternative child mental health problems (e.g., emotional problems) or parental mental health problems (e.g., parental depression). Instead, our findings suggest that the Incredible Years parenting program has a robust impact on the aspects of family well-being that it is primarily designed to impact: several aspects of parenting and children's conduct and ADHD problems.

Acknowledgements

This study was funded by the National Institute of Health Research (NIHR) Public Health Research Programme. The views expressed in this paper are those of the authors and not necessarily those of the National Health Service, the NIHR or the Department of Health.

SL received salary support from the UK National Institute for Health Research Biomedical Research Centre at South London and Maudsley, NHS Foundation Trust, and King's College London.

We gratefully acknowledge the principal investigators of the original trials who shared their data. We thank project advisors George Howe and Anna Goodman for their valuable feedback.

References

Note. References marked with an asterisk indicate studies included in the IPD meta-analysis.

Abidin, R.R. (1995). *Parenting Stress Index* (3rd ed). Odessa, FL: Psychological Assessment Resources.

Achenbach, T.M. (1991). *Manual for the Child Behavior Checklist/4-18 and 1991 profile*. Burlington, VT: Department of Psychiatry, University of Vermont.

Arnold, D.S., O'Leary, S.G., Wolff, L.S., & Acker, M.M. (1993). The Parenting Scale: a measure of dysfunctional parenting in discipline situations. *Psychological Assessment, 5*, 137–144.

*Axberg, U., & Broberg, A.G. (2012). Evaluation of "The Incredible Years" in Sweden: The transferability of an American parent-training program to Sweden. *Scandinavian Journal of Psychology, 53*, 224–232.

*Azevedo, A., Seabra-Santos, M.J., Gaspar, M.F., & Homem, T.C. (2013). The Incredible Years Basic Parent Training for Portuguese preschoolers with AD/HD behaviours: does it make a difference? *Child Youth Care Forum, 42*, 403–424.

Barlow, J., Smailagic, N., Huband, N., Roloff, V., & Bennett, C. (2014). Group-based parent training programmes for improving parental psychosocial health. *Cochrane Database of Systematic Reviews.*

Beck, A.T., Ward, C.H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*, 561–571.

Bonell, C., Jamal, F., Melendez-Torres, G.J., & Cummins, S. (2015). 'Dark logic': Theorising the harmful consequences of public health interventions. *Journal of Epidemiology and Community Health, 69*, 95–98.

Conduct Problems Prevention Research Group (1996). *Parenting Practices Inventory.* Unpublished document. Seattle, WA: University of Washington.

Cybulski, L., Mayo-Wilson, E., & Grant, S. (2016). Improving transparency and reproducibility through registration: The status of intervention trials published in clinical psychology journals. *Journal of Consulting and Clinical Psychology*. Advance online publication.

Derogatis, L.R., & Spencer, P.M. (1982). *The Brief Symptom Inventory: Administration, Scoring and Procedures Manual 1*. Baltimore, MD: Clinical Psychometrics Research.

Dwan, K., Altman, D.G., Arnaiz, J.A., Bloom, J., Chan, A.W., Cronin, E., ... & Ghersi, D. (2008). Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PloS ONE, 3*, e3081.

Eyberg, S.M., & Ross, A.W. (1978). Assessment of child behavior problems: The validation of a new inventory. *Journal of Clinical Child Psychology, 7*, 113–116.

Furlong, M., & McGilloway, S. (2012). The Incredible Years Parenting program in Ireland: A qualitative analysis of the experience of disadvantaged parents. *Clinical Child Psychology and Psychiatry, 17*,616–630.

Furlong, M., McGilloway, S., Bywater, T., Hutchings, J., Smith, S.M., & Donnelly, M. (2012). Behavioral and cognitive-behavioural group-based parenting interventions for early-onset conduct problems in children age 3-12 years. *Cochrane Database of Systematic Reviews*.

*Gardner, F., Burton, J., & Klimes, I. (2006). Randomised controlled trial of a parenting intervention in the voluntary sector for reducing child conduct problems: outcomes and mechanisms of change. *Journal of Child Psychology and Psychiatry, 47*, 1123–1132.

Ginsburg, G.S., & Schlossberg, M.C. (2002). Family-based treatment of childhood anxiety disorders. *International Review of Psychiatry, 14*, 143-154.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry, 38,* 581–586.

Gross, D., Garvey, C., Julion, W., Fogg, L., Tucker, S., & Mokros, H. (2009). Efficacy of the Chicago Parent Program with low-income African American and Latino parents of young children. *Prevention Science, 10*, 54–65.

Hinshaw, S.P. (1992). Externalizing behavior problems and academic underachievement in childhood and adolescence: Causal relationships and underlying mechanisms. *Psychological Bulletin, 111*, 127–155.

*Hutchings, J., Bywater, T., Daley, D., Gardner, F., Whitaker, C., Jones, K., ... & Edwards, R.T. (2007). Parenting intervention in Sure Start services for children at risk of developing conduct disorder: Pragmatic randomised controlled trial. *British Medical Journal*, *334*, 678–682.

* Hutchings, J., Griffith, N., Bywater, T. and Williams, M.E. (2017). Evaluating the Incredible Years Toddler Parenting Programme with parents of toddlers in disadvantaged (Flying Start) areas of Wales. *Child: Care, Health and Development, 43,* 104–113.

Jones, K., Daley, D., Hutchings, J., Bywater, T., & Eames, C. (2007). Efficacy of the Incredible Years Basic Parent Training Programme as an early intervention for children with conduct disorder and ADHD. *Child: Care, Health and Development, 33*, 749–756.

Johnston, C., & Marsh, E.J. (1989) A measure of parenting satisfaction and efficacy. *Journal of Clinical Child Psychology, 18,* 167–175.

Kaehler, L.A., Jacobs, M., & Jones, D.J. (2016). Distilling Common History and Practice Elements to Inform Dissemination: Hanf-Model BPT Programs as an Example. *Clinical Child and Family Psychology Review*, *19*, 236–258.

*Larsson, B., &. Fossum, S., Clifford, G., Drugli, M.B., Handegård, B.H., & Mørch, W.T. (2009). Treatment of oppositional defiant and conduct problems in young Norwegian children: results of a randomized controlled trial. *European Child & Adolescent Psychiatry,18*, 42–52.

Leijten, P., Melendez-Torres, G.J., Knerr, W., & Gardner, F. (2016). Transported versus homegrown parenting interventions for reducing disruptive child behavior: A multilevel meta-regression study. *Journal of the American Academy of Child and Adolescent Psychiatry, 55*, 233–243.

*Leijten, P., Raaijmakers, M.A.J., Orobio de Castro, B., van den Ban, E., & Matthys, W. (2017). Effectiveness of the Incredible Years parenting intervention for families with socioeconomically disadvantaged and ethnic minority backgrounds. *Journal of Clinical Child and Adolescent Psychology, 46*, 59–73.

Levy, F., Hawes, D., & Johns, A. (2015). Externalizing and internalizing comorbidity. In: T.P. Beauchaine & S.P. Hinshaw (eds.) *The Oxford handbook of externalizing spectrum disorders* (pp. 443–460). Oxford: Oxford University Press.

Littell, J.H., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis*. Oxford, Oxford University Press.

*Morpeth, L., Blower, S., Tobin, K., Taylor, R. S., Bywater, T., Edwards, R. T., ... & Berry, V. (2017). The effectiveness of the Incredible Years pre-school parenting programme in the United Kingdom: a pragmatic randomised controlled trial. *Child Care in Practice*. Advance online publication.

*McGilloway, S., Mhaille, G.N., Bywater, T., Furlong, M., Leckey, Y., Kelly, P., … & Donnelly, M. (2012). A parenting intervention for childhood behavioral problems: A randomised controlled trial in disadvantaged community-based settings. *Journal of Consulting and Clinical Psychology, 80,* 116–127.

McLaughlin, K.A., Gadermann, A.M., Hwang, I., Sampson, N.A., Al-Hamzawi, A., Andrade, L.H., ... & Caldas-de-Almeida, J.M. (2012). Parent psychopathology and offspring mental disorders: results from the WHO World Mental Health Surveys. *The British Journal of Psychiatry, 200,* 290–299.

Menting, A.T.A., Orobio de Castro, B., & Matthys, W. (2013). Effectiveness of the Incredible Years parent training to modify disruptive and prosocial child behavior: A meta-analytic review. *Clinical Psychology Review, 33*, 901–913. doi:10.1016/j.cpr.2013.07.006

*Menting, A.T.A., Orobio de Castro, B., Wijngaards-de Meij, L.D.N.V., & Matthys, W. (2014). A trial of parent training for mothers being released from incarceration and their children. *Journal of Clinical Child & Adolescent Psychology, 43*, 381–396.

National Collaborating Centre for Mental Health (2009). *Attention deficit hyperactivity disorder: diagnosis and management of ADHD in children, young people and adults*. London: The Royal College of Psychiatrists and The British Psychological Society.

Patterson, J., Barlow, J., Mockford, C., Klimes, I., Pyper, C., & Stewart-Brown, S. (2002). Improving mental health through parenting programmes: block randomised controlled trial. *Archives of Disease in Childhood, 87*, 472–477.

Robling, M., Bekkers, M.J., Bell, K., Butler, C.C., Cannings-John, R., Channon, S., ... & Kenkre, J. (2016). Effectiveness of a nurse-led intensive home-visitation programme for first-time teenage mothers (Building Blocks): a pragmatic randomised controlled trial. *The Lancet, 387,* 146–155.

Scott, J., & Young, A.H. (2016). Psychotherapies should be assessed for both benefit and harm. *The British Journal of Psychiatry*, *208*, 208–209.

Scott, S. (2012). Parenting quality and children's mental health: biological mechanisms and psychological interventions. *Current Opinion in Psychiatry*, *25*, 301–306.

*Scott, S., O'Connor, T. G., Futh, A., Matias, C., Price, J., & Doolan, M. (2010a). Impact of a parenting program in a high-risk, multi-ethnic community: The PALS trial. *Journal of Child Psychology and Psychiatry*, *51*, 1331−1341.

*Scott, S., Sylva, K., Beckett, C., Doolan, M., Kallitsoglou, A., … & Ford, T. (2013). *Which type of parenting programme best improves child behaviour and reading? The Helping Children Achieve trial. Follow-up of the helping children achieve trial*. Nuffield Foundation, London.

*Scott, S., Sylva, K., Doolan, M., Price, J., Jacobs, B., Crook, C., & Landau, S. (2010b). Randomised controlled trial of parent groups for child antisocial behaviour targeting multiple risk factors: the SPOKES project. *Journal of Child Psychology and Psychiatry*, *51*, 48−57.

*Scott, S., Spender, Q., Doolan, M., Jacobs, B., & Aspland, H. (2001). Multicentre controlled trial of parenting groups for childhood antisocial behaviour in clinical practice. *British Medical Journal*, *323*, 194–198.

Shelton, K.K., Frick, P.J, & Wootton, J. (1996). Assessment of parenting practices in families of elementary school-age children. *Journal of Clinical Child Psychology*, *25*, 317–329.

Sonuga-Barke, E.J., Brandeis, D., Cortese, S., Daley, D., Ferrin, M., Holtmann, M., ... & Dittmann, R.W. (2013). Nonpharmacological interventions for ADHD: Systematic review and meta-analyses of randomized controlled trials of dietary and psychological treatments. *American Journal of Psychiatry*, *170*, 275–289.

Stewart, L.A., Clarke, M., Rovers, M., Riley, R.D., Simmonds, M., Stewart, G., & Tierney, J. F. (2015). Preferred reporting items for a systematic review and meta-analysis of individual participant data: The PRISMA-IPD statement. *JAMA*, *313*, 1657–1665.

Taylor, E., Schachar, R., Thorley, G., & Wieselberg, M. (1986). Conduct disorder and hyperactivity: I. Separation of hyperactivity and antisocial conduct in British child psychiatric patients. *The British Journal of Psychiatry*, *149*, 760–767.

Webster-Stratton, C., & Herman, K.C. (2008). The impact of parent behavior-management training on child depressive symptoms. *Journal of Counseling Psychology, 55*, 473–484.

Webster-Stratton, C., & Reid, M.J. (2010). The Incredible Years Parents, Teachers, and Children Training Series: A multifaceted treatment approach for children with conduct disorders. In J.R. Weisz & A.E. Kazdin, *Evidence-based psychotherapies for children and adolescents* (pp. 194–210). New York, NY: Guilford Press.

Weisz, J.R., & Kazdin, A.E. (2010). Evidence-based psychotherapies for children and adolescents (2nd ed). New York, NY: Guildford Press.

Williams, P., & Goldberg, D.P. (1988). *A user's guide to the General Health Questionnaire.* Berkshire: NFER, Nelson.

Table 1. *Trial Characteristics.*

Trial	Lead author (year)	Country	Setting	Children screened for conduct problems	Baseline ECBI scores (*M*)	*n*	Child age (*M*)	% low income	% ethnic minority
#1	Larsson (2009)	Norway	Outpatient psychiatric clinics	Yes	158.04	75	3–8 (6.58)	25	1
#2	Axberg (2012)	Sweden	Outpatient psychiatric clinics	Yes	155.01	62	3–8 (5.97)	41	0
#3	Azevedo (2013)	Portugal	University clinics	Yes	127.61	124	3–6 (4.66)	0	0
#4	McGilloway (2012)	Ireland	Community services	Yes	158.54	149	2–7 (4.84)	47	6
#5	Menting (2014)	Netherlands	Community services	No	109.66	99	1–11 (6.30)	93	78
#6	Leijten (2017)	Netherlands	Outpatient psychiatric clinics & schools	Yes & No	124.24	156	2–8 (5.59)	74	65
#7	Hutchings (2007)	Wales	Community services	Yes	145.06	153	3–4 (3.84)	79	1
#8	Hutchings (2017)	Wales	Community services	No	N.A.	103	0–2 (1.85)	56	0
#9	Morpeth (2017)	England	Community services	Yes	143.08	161	2–4 (3.68)	63	52
#10	Scott (2010b)	England	Schools	Yes	128.42	112	4–6 (5.21)	44	40

#11	Scott (2010a)	England	Schools	No	104.78	174	4–6 (5.50)	44	75
#12	Scott (2013)	England	Schools	Yes	136.13	214	3–7 (6.07)	80	19
#13	Gardner (2006)	England	Community services	Yes	161.47	76	2–9 (5.93)	64	2
#14	Scott (2001)	England	Outpatient psychiatric clinics	Yes	162.52	141	2–10 (5.67)	58	15

Note. Numbers of participants may differ from the trials' original numbers due to exclusion of alternative intervention conditions and inclusion of one child per family.

Table 2. *Estimated Effects of Incredible Years.*

			Control				Incredible Years				Univariate Analysis				(A) Multivariate Analysis			
			Pretest		Posttest		Pretest		Posttest									
	k	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	β	*p*	95% CI		I^2 (%)	β	*p*	95% CI	
Positive parenting[a]																			
Praise	6	630	5.4	1.2	4.8	1.1	5.4	1.2	5.2	1.2	.26	.045*	.01	.51	57.9	.28	.001***	.12	.44
Tangible rewards	6	625	3.3	1.3	3.4	1.3	3.3	1.2	3.6	1.2	.15	.347	-.16	.45	0.0	.17	.077	-.02	.35
Monitoring	9	1088	5.2	1.7	5.3	1.7	5.3	1.7	5.4	1.6	.05	.434	-.08	.18	0.0	.03	.625	-.10	.16
Negative parenting[a]																			
Corporal punishment	10	1393	2.2	1.4	2.4	1.6	2.1	1.4	2.0	1.4	-.22	.004**	-.42	-.01	37.0	-.19	.005**	-.32	-.05
Threatening	9	999	3.6	1.5	3.2	1.5	3.5	1.6	2.8	1.5	-.21	.007**	-.36	-.06	0.0	-.20	.003**	-.34	-.07
Laxness	9	978	3.3	1.3	3.3	1.2	3.3	1.3	3.1	1.2	-.15	.174	-.37	.07	43.0	-.15	.045*	-.29	.00
Shouting	9	967	3.3	1.5	3.0	1.3	3.1	1.3	2.7	1.3	-.31	.041*	-.61	-.01	21.0	-.22	.001**	-.35	-.08
Child mental health																			
Conduct problems[b]	13	1622	135.5	37.0	125.5	37.9	139.4	37.0	116.2	34.7	-.35	.000***	-.51	-.19	42.5	-.34	.000***	-.45	-.23
ADHD symptoms[c]	11	1532	5.8	2.7	5.9	0.1	5.9	2.7	5.3	0.1	-.30	.000***	-.44	-.17	0.0	-.28	.000***	-.41	-.15
Emotional problems[c]	10	1340	3.2	2.4	2.7	0.1	3.4	2.7	2.7	0.1	-.06	.303	-.18	.06	0.0	-.01	.933	-.13	.11
Parental mental health																			
Depression [d]	11	1395	10.1	9.7	8.7	9.0	12.2	10.9	8.7	9.1	-.08	.095	-.17	.01	0.0	-.08	.158	-.19	.03

Parenting stress[e]	5	542	89.0	28.4	82.9	34.9	92.1	28.4	80.5	33.6	-.18	.164	-.44	.07	0.0	-.08	.280	-.21	.06
Self-efficacy[f]	4	417	54.1	7.6	59.4	13.2	54.0	7.6	55.7	14.1	-.32	.165	-.77	.13	6.8	-.31	.083	-.66	.04

Note. *k* = number of trials contributing data; *n* = number of participants; **p* <.05; ***p* <.01; ****p* <.001; I^2 = heterogeneity in effects across studies: <30% indicates low levels of heterogeneity and >70 % indicates considerable heterogeneity; Possible ranges in scores [a]1–7; [b]36–252; [c]0–10; [d]0–63; [e]36–216; [f]16–96.

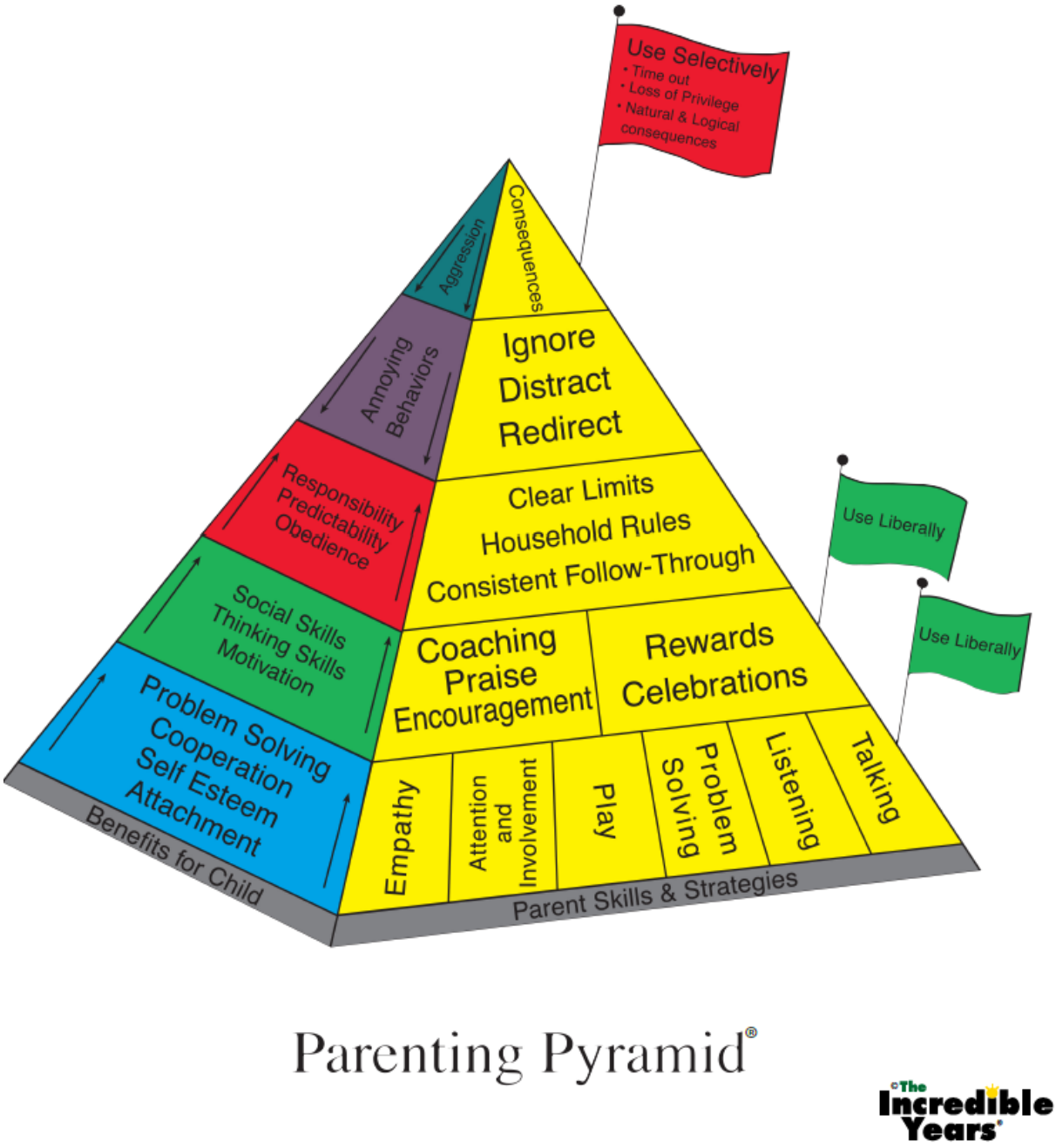

Figure 1. Core Components of the Incredible Years Parent Program.

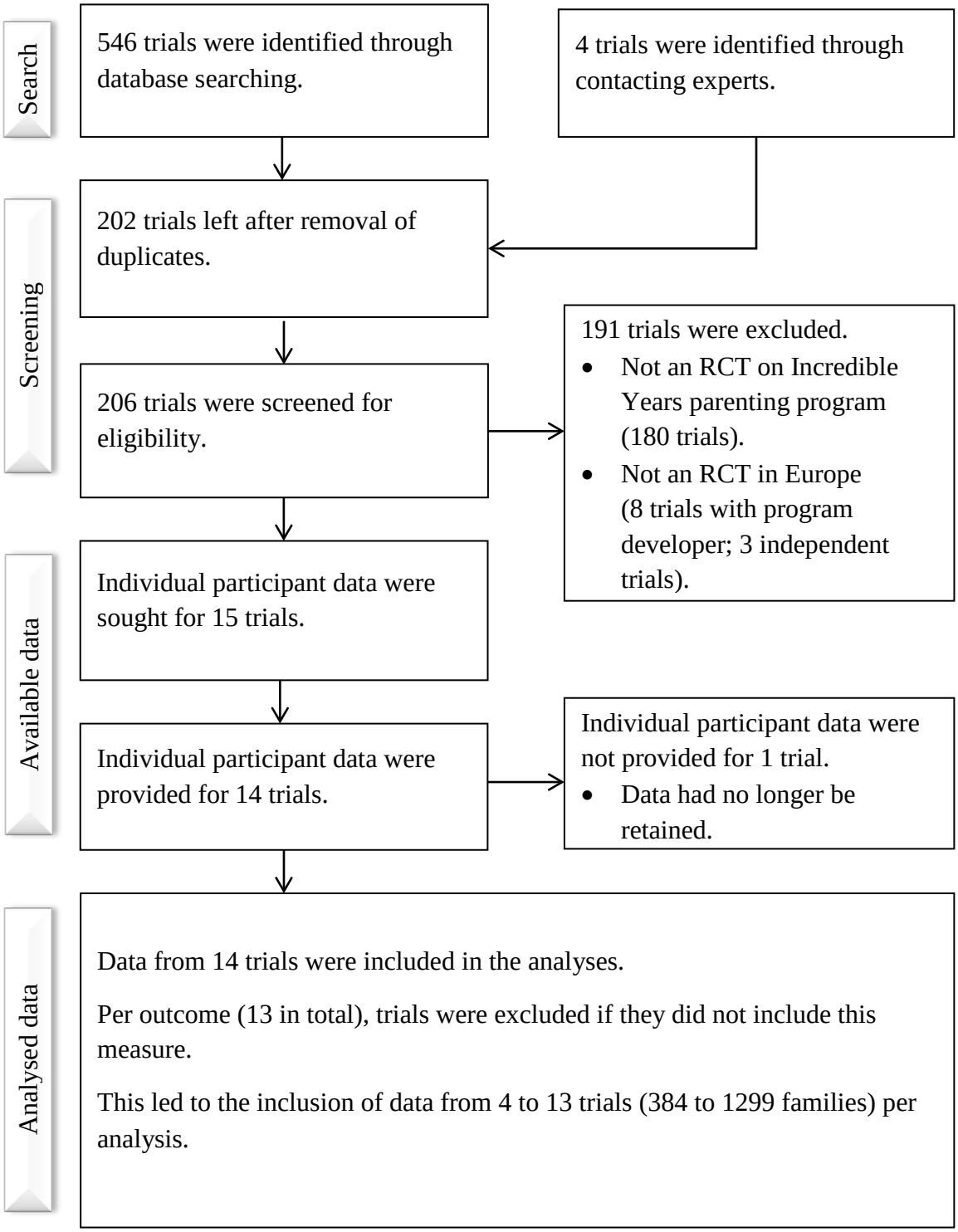

Figure 2. Flow Diagram of Search Results.

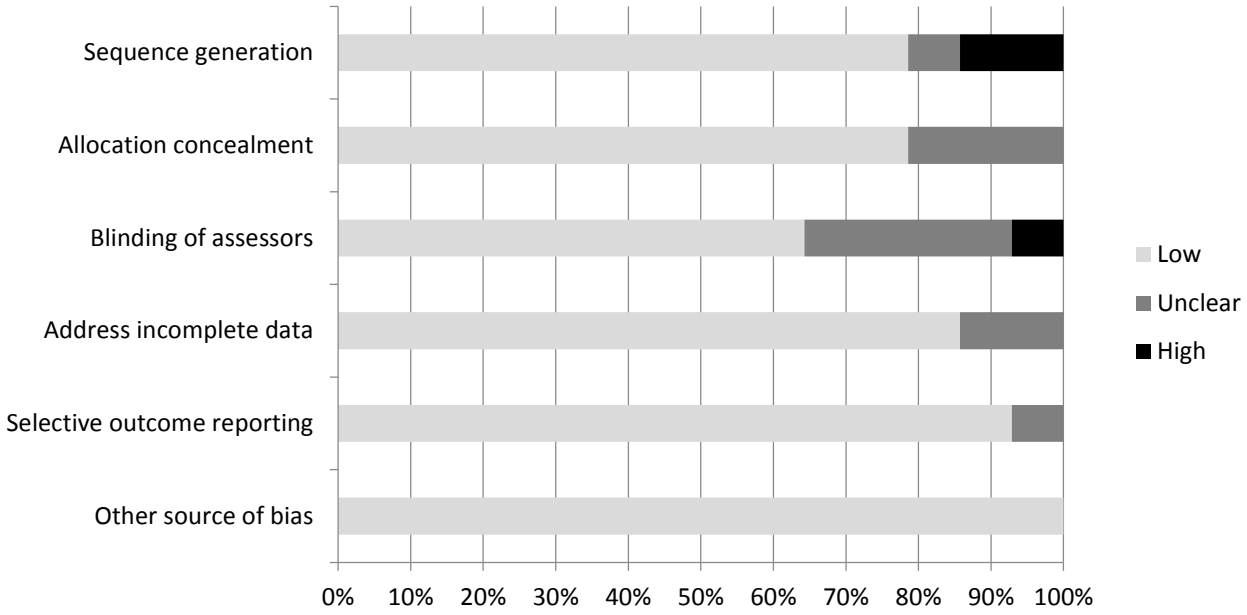

Figure 3. Risk of Bias Within Trials.

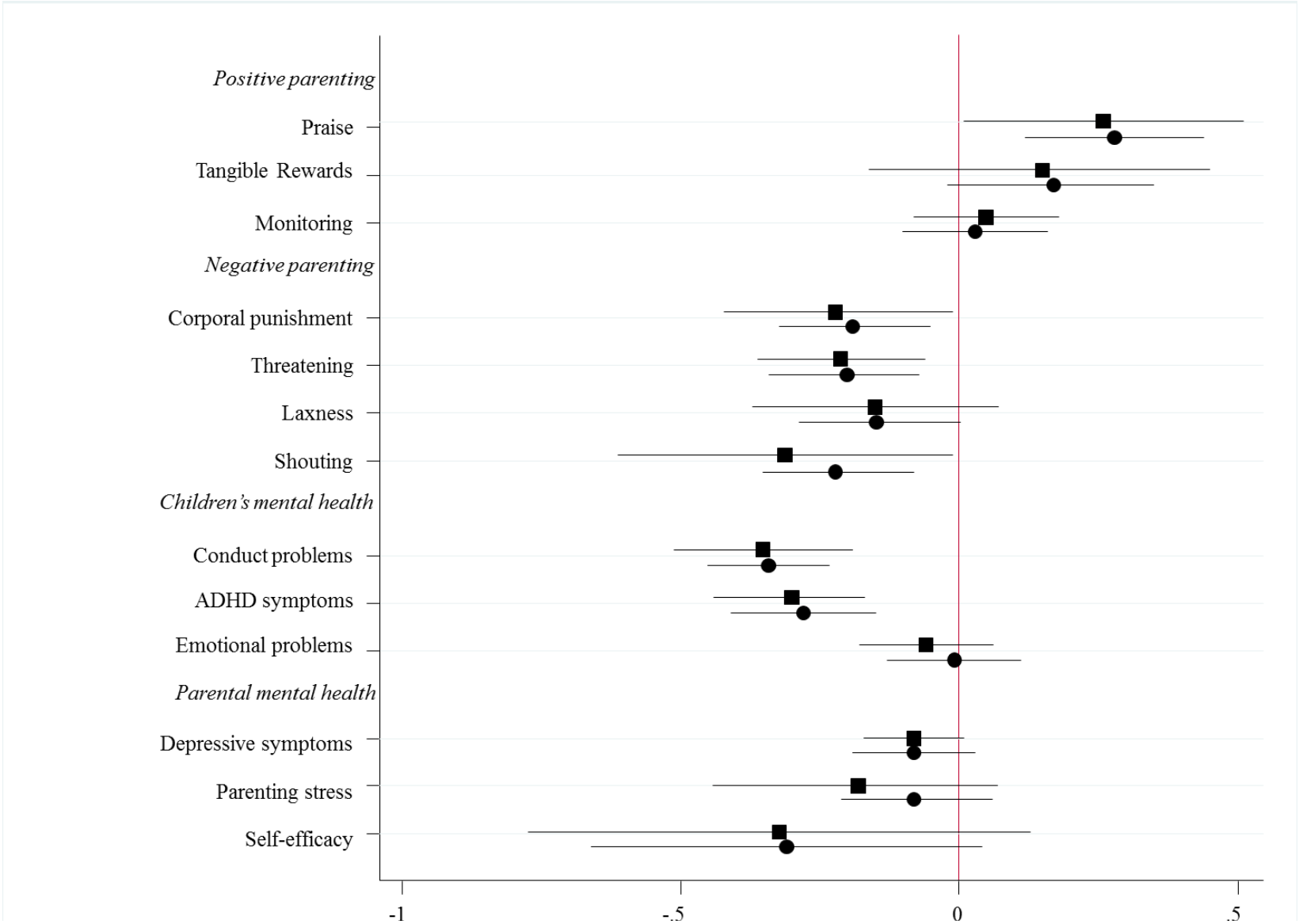

Figure 4. Graphical Summary of the Results.

Note. X-axis shows the effect size (β): group differences expressed in baseline standard deviations; For all outcomes except Praise, Tangible Rewards, Monitoring, and Self-efficacy, negative regression coefficient reflect benefits of the Incredible Years program; ■ = Univariate analysis; ● = Multivariate analysis;— = 95% confidence interval.

Key points

- This is the first individual participant data meta-analysis of a parenting program.
- Incredible Years improved parent-reported conduct problems and ADHD, and several aspects of parenting.
- Incredible Years did not improve children's emotional problems or parental mental health.
- There were no signs of harm on any of the target outcomes.